PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-029496
(43)Date of publication of application : 28.01.2000

(51)Int.Cl. G10L 15/28
G10L 15/20
G10L 15/18

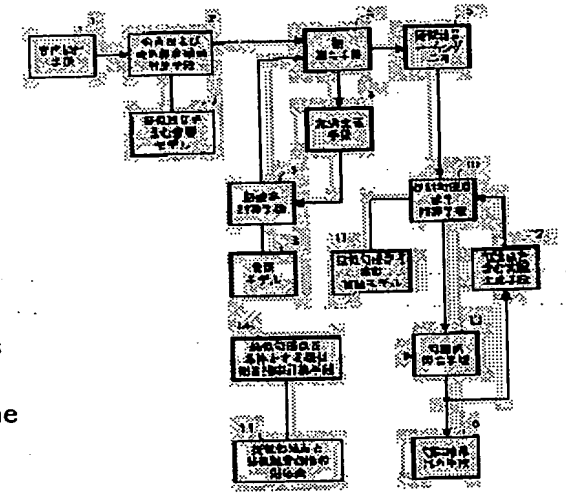
(21)Application number : 11-132117
(22)Date of filing : 13.05.1999

(71)Applicant : INTERNATL BUSINESS MACH CORP <IBM>
(72)Inventor : DONALD T TAN
SHAO CHIN CHU
LEE CHIN SHEN

(30)Priority
Priority number : 98 98108367 Priority date : 13.05.1998 Priority country : CN

(54) APPARATUS FOR AUTOMATICALLY GENERATING PUNCTUATION IN CONTINUOUS RECOGNITION AND METHOD THEREFOR

(57)Abstract:
PROBLEM TO BE SOLVED: To provide an apparatus for automatically generating punctuation in a continuous voice recognition and a method therefor.
SOLUTION: This apparatus automatically generates the punctuation with a continuous voice recognition system including means 1, 2, 3, 5 for recognizing user's voice and converting the user's voice into words. In such a case, the means 1, 2, 3, 5 for recognizing the user's voice are used also for recognizing the pseudo noise in the user's voice. Further, the apparatus includes a means 9 for marking the pseudo noise in the output results of the means 1, 2, 3, 5 for recognizing the user's voice and means 10, 14, 13 for generating the punctuation by finding out the pseudo punctuation of the highest probability in the position of the pseudo noise marked by the means 9 for marking the pseudo noise in accordance with the language model including the pseudo punctuation.



LEGAL STATUS
[Date of request for examination] 27.12.1999
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]

[Patent number]	3282075
[Date of registration]	01.03.2002
[Number of appeal against examiner's decision of rejection]	
[Date of requesting appeal against examiner's decision of rejection]	
[Date of extinction of right]	01.03.2005

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2000－29496
（P2000－29496A）
(43)公開日　平成12年１月28日(2000.1.28)

(51)Int.Cl.⁷	識別記号	ＦＩ		ﾃｰﾏｺｰﾄﾞ（参考）
Ｇ１０Ｌ	15/28	Ｇ１０Ｌ	3/00	５６１Ｈ
	15/20			５３１Ｐ
	15/18			５３７Ｇ

審査請求　未請求　請求項の数６　ＯＬ　（全 10 頁）

(21)出願番号　特願平11－132117

(22)出願日　平成11年５月13日(1999.5.13)

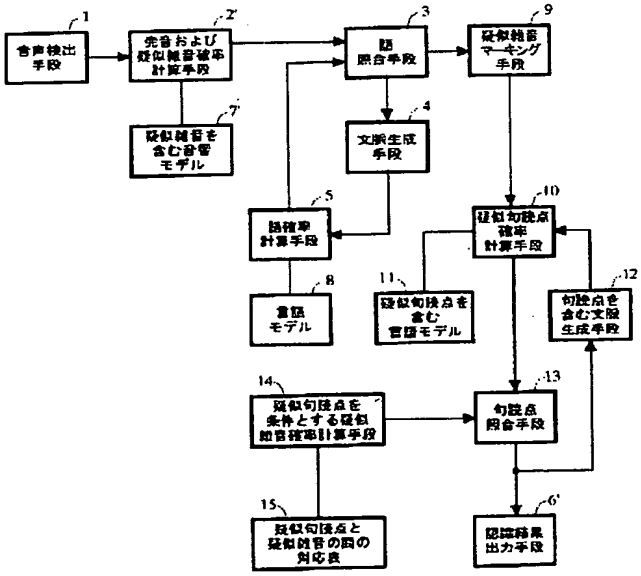
(31)優先権主張番号　９８１０８３６７．６
(32)優先日　平成10年５月13日(1998.5.13)
(33)優先権主張国　中国（ＣＮ）

(71)出願人　390009531
インターナショナル・ビジネス・マシーンズ・コーポレイション
ＩＮＴＥＲＮＡＴＩＯＮＡＬ　ＢＵＳＩＮＥＳＳ　ＭＡＳＣＨＩＮＥＳ　ＣＯＲＰＯＲＡＴＩＯＮ
アメリカ合衆国10504、ニューヨーク州　アーモンク　（番地なし）
(74)代理人　100086243
弁理士　坂口　博　（外１名）

最終頁に続く

(54)【発明の名称】　連続音声認識において句読点を自動的に生成する装置および方法

(57)【要約】
【課題】　連続音声認識で句読点を自動的に生成する装置および方法を提供すること。
【解決手段】　ユーザ音声を認識し、ユーザ音声を語に変換する手段（１、２、３、５）を含む、連続音声認識システムで句読点を自動的に生成する装置であって、ユーザ音声を認識する手段（１、２、３、５）はユーザ音声中の擬似雑音を認識するためにも使用され、さらにユーザ音声を認識する手段（１、２、３、５）の出力結果において擬似雑音をマークする手段（９）と、擬似句読点を含む言語モデルに基づいて擬似雑音をマークする手段（９）によってマークされた擬似雑音の位置において最も可能性の高い擬似句読点を見つけ出すことによって、句読点を生成する手段（１０、１４、１３）とを含むことを特徴とする。



1

【特許請求の範囲】

【請求項１】連続音声認識システムで句読点を自動的に生成する装置であって、

ユーザ音声を認識して語に変換し、さらに前記ユーザ音声中の擬似雑音も認識する手段（１、２、３、５）と、

前記認識する手段（１、２、３、５）の出力結果における擬似音声をマークする手段（９）と、

擬似句読点を含む言語モデルに基づいて、前記マークする手段（９）によってマークされた擬似雑音の位置において最も可能性の高い擬似句読点を見つけ出すことによって句読点を生成する手段（１０、１４、１３）とを含む装置。

【請求項２】句読点を生成する前記手段が、

擬似句読点を含む言語モデル内の各擬似句読点について、前記出力結果に前記擬似句読点が現れる確率を計算する手段（１０）と、

特定の擬似雑音が特定の擬似句読点の位置に現れる確率を計算する手段（１４）と、

計算された前記確率に基づいて、前記マークされた擬似雑音の位置において最も可能性の高い擬似句読点を見つけ出し、前記最も可能性の高い擬似句読点に対応する句読点を生成する手段（１３）とを含む請求項１記載の装置。

【請求項３】ユーザ音声を認識し、前記ユーザ音声を語に変換する手段（１、２、３、５）を含む、連続音声認識システムで句読点を自動的に生成する装置であって、

口述中のユーザの操作に応答して、前記手段（１、２、３、５）の出力結果中の位置を指示する位置指示信号を生成する手段と、

擬似句読点を含む言語モデル内の各擬似句読点について、前記擬似句読点が前記出力結果に現れる確率を計算する手段（１０）と、

計算された前記確率に基づいて、前記位置指示信号によって指示された位置において最も可能性の高い擬似句読点を見つけ出し、前記最も可能性の高い擬似句読点に対応する句読点を生成する手段（１３）とを含む装置。

【請求項４】連続音声認識システムで句読点を自動的に生成する方法であって、

ユーザ音声を認識して語に変換し、さらに前記ユーザ音声中の擬似雑音も認識するステップと、

前記ステップの出力結果において擬似雑音をマークするステップと、

擬似句読点を含む言語モデルに基づいて、擬似雑音をマークする前記ステップでマークされた前記擬似雑音の位置において最も可能性の高い擬似句読点を見つけ出すことによって句読点を生成するステップと、を含む方法。

【請求項５】句読点を生成する前記ステップが、

擬似句読点を含む言語モデル内の各擬似句読点について、前記出力結果に前記擬似句読点が現れる確率を計算するステップと、

2

特定の擬似雑音が特定の擬似句読点の位置に現れる確率を計算するステップと、

計算された前記確率に基づいて、前記マークされた擬似雑音の位置において最も可能性の高い擬似句読点を見つけ出し、前記最も可能性の高い擬似句読点に対応する句読点を生成するステップとを含む請求項４記載の方法。

【請求項６】ユーザ音声を認識し、前記ユーザ音声を語に変換するステップを含む、連続音声認識システムで句読点を自動的に生成する方法であって、

口述中のユーザの操作に応答して、前記ステップの出力結果中の位置を指示する位置指示信号を生成するステップと、

擬似句読点を含む言語モデル内の各擬似句読点について、前記擬似句読点が前記出力結果に現れる確率を計算するステップと、

計算された前記確率に基づいて、前記位置指示信号によって指示された位置において最も可能性の高い擬似句読点を見つけ出し、前記最も可能性の高い擬似句読点に対応する句読点を生成するステップとをさらに含むことを特徴とする方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、連続音声認識技術に関し、さらに詳しくは、連続音声認識において句読点を自動的に生成する装置および方法に関する。

【０００２】

【従来の技術】一般的な音声認識システムは、図１のように示すことができる。システムは一般的に音響モデル７および言語モデル８を含む。音響モデル７は、認識される言語で一般的に使用される語の発音を含む。そのような語の発音は、大抵の人々がこの語を読むときの発音から、統計的な方法を用いることによって要約され、その語に特徴的な一般的発音を表す。言語モデル８は、認識される言語で一般的に使用される語が利用される方法を含む。

【０００３】図１に示す連続音声認識システムの動作手順は次の通りである。音声検出手段１が例えばユーザの音声を収集し、言語を音声サンプルで表現し、音声サンプルを発音確率計算手段２に送る。音響モデル７内の全ての発音について、発音確率計算手段２は、それが音声サンプルと同じであるかどうかの確率推定値を与える。語確率計算手段５は、大量の言語材料から要約された言語規則に従って、言語モデル８内の語について、それが現在の文脈に現れるかどうかの確率推定値を与える。語照合手段３は、発音確率計算手段２によって計算された確率値を、語確率計算手段５によって計算された確率値と組み合わせることによって結合確率（音声サンプルをこの語と認識する可能性を表す）を計算し、最大の結合確率値を持つ語を音声認識の結果として選ぶ。文脈生成手段４は、上述の認識結果を使用することによって現在

の文脈を、次の音声サンプルの認識で使用するように変更する。語出力手段６は、認識された語を出力する。

【０００４】上述の連続認識手順は、文字、単語、または語句単位で実行することができる。したがって、以下では文字、単語、または語句を語と呼ぶ。

【０００５】認識された結果に句読点を付けるために、現在の連続音声認識システムは、口述中に句読点を声に出して言う必要があり、そうするとそれを認識する。例えば、「Ｈｅｌｌｏｗ！　Ｗｏｒｌｄ．」を完全に認識するためには、話者は、「Ｈｅｌｌｏｗ感嘆符ｗｏｒｌｄ終止符」と言わなければならない。つまり、現在の音声認識システムでは、話者が句読点を音声に変換しなければならず（つまり、句読点を声に出して言わなければならず）、そうすると句読点は対応する句読点として音声認識システムに認識される。したがって、言語モデルは句読点を含む必要があり、つまり言語モデル８は、全ての句読点についてそれが現在の文脈に現れる句読点であるかどうかの推定確率値を与えることができる。

【０００６】しかし、上述の音声認識システムを使用することによって自然なスピーチ活動（例えば会議、ラジオ放送、およびＴＶ番組など）を文字化するとき、人々が句読点を声に出して言うことを期待することはできない。さらに、口述中に句読点を声に出して言うことは、極めて不自然である。そうするように要求された場合でさえも、人々は、話し中または記事を読み上げるときに、句読点を声に出して言うことをしばしば忘れる。さらに、全ての文章が心から直接生じる自然発生的なスピーチの口述では、大抵の人々にとって、使用すべき句読点を正確に決定し、流暢さを損ねることなく全ての句読点を正確に声に出して言うことは、非常に困難である。これは、日常的な音声言語で句読点が使用されることは、皆無ではないとしても、まれであるという事実の結果である。

【０００７】

【発明が解決しようとする課題】したがって、連続音声認識において容易に使用され、スピーチ中に句読点を声に出して言うことを必要とせず、したがってユーザの通常のスピーチに影響を及ぼさない、句読点を自動的に生成する装置および方法が緊急に必要である。

【０００８】本発明の第一の目的は、連続音声認識で句読点を自動的に生成する装置を提供することである。

【０００９】本発明の第二の目的は、連続音声認識で句読点を自動的に生成する方法を提供することである。

【００１０】

【課題を解決する手段】第一の目的を達成するために、本発明は、ユーザの音声を語として認識する音声認識手段を含む、連続音声認識で句読点を自動的に生成するための装置を提供する。この音声認識手段はユーザの音声中の疑似雑音をも認識し、さらに、音声認識手段の出力結果における疑似雑音をマークする疑似雑音マーキング手段と、疑似句読点を含む言語モデルに基づいて、疑似雑音マーキング手段によってマークされた疑似雑音の全ての位置において最も可能性の高い疑似句読点を見つけ出すことによって、最も可能性の高い疑似句読点に対応する句読点を生成する句読点生成手段とを含む。

【００１１】本発明はさらに、ユーザの音声を語として認識するための音声認識手段と、口述中のユーザの操作に応答して、音声認識手段の出力結果における位置を指示する位置指示信号を生成するための句読点位置指示手段と、言語モデルに含まれる全ての疑似句読点について、それが音声認識手段の出力結果に現れる確率推定値を与える疑似句読点確率計算手段と、前記疑似句読点確率計算手段によって計算された確率推定値に基づき、位置指示信号によって指示された位置の疑似句読点を見つけ出すことによって、疑似句読点に対応する句読点を生成する句読点照合手段とを含む、連続音声認識において句読点を自動的に生成するための装置を提供する。

【００１２】第二の目的を達成するために、本発明は、ユーザの音声を語として認識し、ユーザ音声中の疑似雑音をも認識する音声認識段階と、音声認識段階の結果出力において疑似雑音をマークする疑似雑音マーキング段階と、疑似句読点を含む言語モデルに基づいて、疑似雑音マーキング段階でマークされた疑似雑音の全ての位置において最も可能性の高い疑似句読点を見つけ出すことによって、最も可能性の高い疑似句読点に対応する句読点を生成する句読点生成段階とを含む、連続音声認識において句読点を自動的に生成する方法を提供する。

【００１３】本発明はさらに、ユーザの音声を語として認識する音声認識段階と、口述中のユーザの操作に応答して、音声認識段階の出力結果における位置を指示する位置指示信号を生成する句読点位置指示段階と、言語モデルに含まれる全ての疑似句読点について、それが前記音声認識段階の出力結果に現れる確率推定値を与える疑似句読点確率計算段階と、前記疑似句読点確率計算段階によって計算された確率推定値に基づき、位置指示信号によって指示される位置に疑似句読点を見つけ出すことによって、疑似句読点に対応する句読点を生成する句読点照合段階とを含む、連続音声認識において句読点を自動的に生成する方法を提供する。

【００１４】本発明の装置および方法では、システムが自動的に句読点を生成することができるので、ユーザが句読点を声に出して言う必要がない。したがって、本発明の装置および方法により、ユーザのスピーチの流暢さが悪影響を受けず、音声認識システムの正確さおよび高速性を向上することができる。

【００１５】

【発明の実施の形態】最初に、本発明の幾つかの概念を紹介する。

【００１６】日常のスピーチにおいて、言語の語に対応する連続音声を出す他に、人々はしばしば、吸息や唇を

5

打つ音など、多少の雑音を出している。これらの雑音は言語の語として認識することはできない。さらに、連続音声と連続音声の間に沈黙が入ることがある。一般的な音声認識システムは、これらの雑音や沈黙を使用せず、これらを除去するだけである。経験から本発明者らは、雑音および沈黙と表記すべき句読点との間に特定の関係があることを発見した。例えば、記事を読み上げているときに、終止符「．」が現れると、人々は習慣的に長時間沈黙し続ける。またコンマ「，」が現れると、彼らはしばしば短時間沈黙し続け、急いで吸息する。「、」が現れるときは、さらに短い時間吸息せずに沈黙し続けるだけである。したがって、本発明の方法では、これらの雑音および沈黙を利用する。

【００１７】さらに、２つの語が間に音や小休止を入れずに滑らかに話されるときに、それらの間に句読点が入ることがある。本発明の方法を実現するために、２つの連続する語の間に「無音」記号を人工的に追加する。本明細書では、雑音、沈黙、および「無音」を擬似雑音と呼ぶ。したがって、２つの語の音の間には必ず擬似雑音がある。

【００１８】全ての擬似雑音は、擬似雑音集合Ｄを構成する。したがって、

Ｄ＝｛「無音」，沈黙，吸息，唇を打つ音，．．．｝

となる。

【００１９】言語には句読点をマークする特定の規則がある。コンピュータによる句読点の自動マーキングの実現を促進するために、句読点を含む莫大な量の音声材料から統計的な方法によって句読点をマークする規則を要約する必要がある。本発明の方法の実現を促進するために、テキストの句読点が表れない場所に、意図的に「無句読点」を追加する。本明細書では、句読点および「無句読点」を擬似句読点と定義する。

【００２０】したがって、２つの語の間には必ず擬似句読点がある。

【００２１】全ての擬似句読点は、擬似句読点集合Ｍを構成する。

【００２２】Ｍ＝｛「無句読点」，「終止符」，「コンマ」，「感嘆符」，「小休止」，．．．｝

となる。

【００２３】句読点の自動生成は、２つの必要な段階を含む。第一段階で、句読点をどこにマークすべきかを決定する。つまり句読点の位置を決定する。第二段階で、どの句読点をマークすべきかを決定する。つまり句読点の種類を決定する。以下で、句読点の位置および種類の決定を自動的に完了できる、より複雑な第一実施形態について説明する。次に、ユーザが句読点の位置を指示しなければならない第二実施形態について説明する。

【００２４】図２は、本発明に係る連続音声認識において句読点を自動的に生成するための装置の第一実施形態の一般的構造の略図を示す。図２で、符号１は音声検出

6

手段を表し、符号２’は発音および擬似雑音確率計算手段を、符号３は語照合手段を、符号４は文脈生成手段を、符号５は語確率計算手段を、符号６’は認識結果出力手段を、符号７’は擬似雑音を含む音響モデルを、符号８は言語モデルを表す。上記構成要素は、図１に示す対応する構成要素と同一または同様の機能を持つ。さらに、符号９は擬似雑音マーキング手段を、符号１０は擬似句読点確率計算手段を、符号１１は擬似句読点を含む言語モデルを、符号１２は句読点を含む文脈生成手段を、符号１３は句読点照合手段を、符号１４は擬似句読点を条件とする擬似雑音確率計算手段を、符号１５は擬似句読点と擬似雑音の間の対応表を表す。

【００２５】図２では、擬似雑音集合Ｄの各要素に対応する音響が、擬似雑音を含む音響モデル７’に追加されている（その機能は図１の音響モデル７と同様である）。したがって、擬似雑音を含む音響モデル７’は、語の発音または擬似雑音のいずれかに対応する。擬似雑音を含む音響モデル７’の各発音または雑音について、発音および擬似雑音確率計算手段２’は、それが音声サンプルに近いかどうかの確率推定値を与える。擬似雑音を含む音響モデルを第一音響モデルＡＭ１と呼び、これは各語の発音だけでなく、各擬似雑音に対応する音響をも含む。

【００２６】擬似句読点集合Ｍの各要素は、擬似句読点を含む言語モデル１１に追加される。いうまでもなく、全ての擬似句読点を、同一モデルとして言語モデル８に追加することができる。様々な実装方式は本発明を限定するものではない。語確率計算手段５は、図１の語確率計算手段５と同じであり、そこで使用される言語モデルを第一言語モデルＬＭ１と呼ぶ。第一言語モデルＬＭ１は、認識される言語で頻繁に使用される全ての語を含む。

【００２７】したがって、図１に示す装置と同様に、検出された音響は、音響検出手段１、発音および擬似雑音確率計算手段（ＡＭ１）２’、語照合手段３、文脈生成手段４、語確率計算手段（ＬＭ１）５、擬似雑音を含む音響モデル７’、および言語モデル８を使用することによって、対応する語または擬似雑音に復号することができる。この復号結果を第一シーケンスと呼ぶ。第一シーケンスにおける「無音」など、他の擬似雑音は、擬似雑音マーキング手段９によってマークされる。

【００２８】（擬似句読点を含む）現在の文脈の場合、擬似句読点確率計算手段１０は、句読点を含む大量の言語材料から要約された言語規則に基づいて、擬似句読点を含む言語モデル１１の擬似句読点が次の句読点であるかどうかの確率推定値を計算する。この装置で使用する言語モデル１１を第二言語モデルＬＭ２と呼ぶ。第二言語モデルを構築する際に、音声材料における全ての句読点を確保した。したがって、第二言語モデルＬＭ２は全ての擬似句読点を含む。例えば、ｃを現在の文脈、ｍを

擬似句読点と仮定すると、ＬＭ２の作用は、Ｐ（ｍ｜ｃ）を計算することである。

【００２９】第二音響モデルＡＭ２を使用する、擬似句読点を条件とする擬似雑音確率計算手段１４は、特定の擬似雑音が特定の擬似句読点位置で現れる確率推定値を出す。第二音響モデルＡＭ２は、統計的方法を使用することによって、大量の言語材料に基づいて構築される。第二音響モデルＡＭ２の構築中に、擬似句読点と擬似雑音の対応する対を見つけ出し、擬似句読点と擬似雑音との間の対応表１５に格納する。擬似句読点を条件とする擬似雑音確率計算手段１４は、擬似句読点と擬似雑音の間の対応表１５に基づいて、条件付き確率Ｐ（ｄ｜ｍ）を計算する。ここでｍは擬似句読点であり、ｄは擬似雑音である。第二音響モデルＡＭ２の特定の構築については、後で詳述する。

【００３０】当然、そのような条件付き確率Ｐ（ｄ｜ｍ）は、大量の言語材料を使用する統計的方法によって予め獲得し、対応する表に格納することができる。句読点を生成する実際の手順では、表を検索することによって、対応する確率値が見つけ出される。つまり、擬似句読点を条件とする擬似雑音確率計算手段は、様々な方法で実装することができるが、それは本発明にいかなる制限も加えない。

【００３１】句読点照合手段１３は、擬似句読点確率計算手段１０によって計算された確率推定値Ｐ（ｍ｜ｃ）と、擬似句読点を条件とする擬似雑音確率計算手段１４によって計算された確率推定値Ｐ（ｄ｜ｍ）とを組み合わせ、擬似句読点を含む言語モデル１１の全ての擬似句読点に関する相関確率Ｐ（ｄ｜ｍ）＊Ｐ（ｍ｜ｃ）（擬似雑音を別の擬似句読点と認識する確率を表す）を計算し、最大相関確率値を持つ擬似句読点を自動生成擬似句読点として選択する。この手順は次のように表すことができる。

M^{ML}＝ａｒｇｍａｘ　ｍ：ＡＭ２（ｄ、ｍ）＊ＬＭ２（ｍ，ｃ）

ここでｍは擬似句読点、ｄは雑音、ｃは文脈であり、かつ

ＡＭ２（ｄ，ｍ）＝Ｐ（ｄ｜ｍ）、

ＬＭ２（ｍ，ｃ）＝Ｐ（ｍ｜ｃ）

である。

【００３２】ｍ＝「無句読点」の場合、文脈状態において、句読点の代わりに、語が来ることを表し、したがって、

Ｐ（「無句読点」｜ｃ）＝ΣＰ（ｗ｜ｃ）

となる。ｗ＝語である。

【００３３】句読点を含む文脈生成手段１２は、上述の生成された句読点を使用して現在の文脈を変更させ、次の擬似雑音を処理する。認識結果出力手段６'は、認識された語および自動的に生成された句読点（または変換された通常の句読点）を出力する。

【００３４】本発明に従って句読点を自動的に生成する装置の第二実施形態として、連続音声認識において句読点を自動的に生成するための別の種類の装置が、上述の第一実施形態から誘導される。著しい相違点は、それが、口述中のユーザの操作に応答して、音声認識手段の出力結果における位置を示す位置指示信号を生成する句読点位置指示手段を含むことにある。位置指示手段は、例えばマウスまたはその他の特殊ハードウェアとすることができる。それはまた、言語モデルに含まれる各擬似句読点について、それが音響認識手段の出力結果に現れる確率推定値を与えるための擬似句読点確率計算手段（１０）と、擬似句読点確率計算手段によって計算された確率推定値に従って、位置指示信号によって指示された位置の擬似句読点を見つけ出し、擬似句読点に対応する句読点を生成するための句読点照合手段をも含む。

【００３５】句読点を自動的に生成するための上述の装置では、擬似雑音は利用されない。したがって、第一音響モデルＡＭ１および第二音響モデルＡＭ２における擬似雑音部は除去され、実装は容易になる。一方、より高い精度を得ることができる。しかし、ユーザにとっては、第一実施形態ほど便利ではない。

【００３６】図３は、本発明に係る連続音声認識で句読点を自動的に生成する方法の第一実施形態の流れ図である。

【００３７】ステップＳ３１で、音声認識手順が開始する。このステップでは、文脈ｃなどの内部変数は全て空にされる。

【００３８】ステップＳ３２で、ユーザが語を読む音声が検出される。ステップＳ３３で、ユーザの音声は、第一音響モデルＡＭ１および第一言語モデルＬＭ１を使用することによって、語または擬似雑音に復号される。例えば、次のような文章

「このリンゴは赤く，緑ではない．」

を読み上げるときに、人々はその中の語だけを読み上げる。したがって、以下のステップのそれぞれを繰返し実行することによって、ユーザの音声は次のような第一シーケンスに復号することができる。

「このリンゴは赤く（吸息）緑ではない（沈黙）」。

【００３９】ステップＳ３４で、上記の第一シーケンスにおける擬似雑音がマークされる。ここで擬似雑音とは、ステップＳ３３で復号されなかった他の擬似雑音を指す。この実施形態では、実装を容易にするために、「無音」マークが２つの連続する語の間に追加される。したがって、次のような第二シーケンスが形成される。

「この（無音）リンゴ（無音）は（無音）赤く（吸息）緑（無音）ではない（沈黙）」。

【００４０】ステップＳ３５で、全ての擬似句読点ｍについて、現在の文脈の場合における条件付き確率Ｐ（ｍ｜ｃ）が計算される。

【００４１】ステップＳ３６で、全ての擬似雑音ｄにつ

9

いて、それぞれの擬似句読点ｍの場合の条件付き確率Ｐ（ｄ｜ｍ）が計算される。代替方法として、各擬似雑音ｄおよび各擬似句読点ｍについて、統計的方法を用いることによって大量の言語材料に基づいて、事前に条件付き確率Ｐ（ｄ｜ｍ）を計算し、表に格納しておき、次に表を検索することによってステップＳ３６を実現することもできる。

【００４２】ステップＳ３７で、Ｐ（ｄ｜ｍ）＊Ｐ（ｍ｜ｃ）を最大にする擬似句読点M^{ML}を見つける。つまり、

M^{ML}＝ａｒｇｍａｘ　ｍ：Ｐ（ｄ｜ｍ）＊Ｐ（ｍ｜ｃ）

を計算する。

【００４３】ステップＳ３５、Ｓ３６、およびＳ３７　＊

ｍ＝「無句読点」の場合、

ＬＭ２（「無句読点」，ｃ）＝Ｐ（「無句読点」｜ｃ）

＝Σｃｏｕｎｔ（ｃ，ｗ）

ｗ≠句読点である。これは、句読点でない語ｗの全てのＰ（ｗ｜ｃ）の合計を示す。

【００４６】ステップＳ３８で、M^{ML}は自動的に生成された擬似句読点として選ばれ、現在の文脈ｃが更新される。したがって、次の第三シーケンスが形成される。

「この（無句読点）リンゴ（無句読点）は（無句読点）赤く（コンマ）緑（無句読点）ではない（終止符）」。

【００４７】ステップＳ３９で、連続音声認識を終了するかどうかが判断される。終了しない場合には、ステップＳ３２にジャンプする。そうでなければ、手順はステップＳ３１０に進む。

【００４８】ステップＳ３１０で、認識された語および自動的に生成された句読点が出力される。このステップで、擬似句読点を実句読点に置換することができる。例えば、次のような結果が出力される。

「このリンゴは赤く，緑ではない．」

【００４９】ステップＳ３１１で手順は終了する。

【００５０】上述の第一、第二、および第三シーケンスは、ユーザが各語を読み上げると同時に、ステップＳ３２からＳ３８までを繰返し実行することによって、徐々に形成されることに留意されたい。つまり、上記手順は実時間で実行される。句読点は、文全体の復号が完了した後だけでなく、実時間で自動的に生成することができる。文脈を形成する語の復号が終了すると、文脈に基づいて句読点を生成することができる。いうまでもなく、音声認識は文単位で実行することができる。ただし、それは、本発明にいかなる制限も加えない。

【００５１】上述の通り、第二音響モデルＡＭ２は、大量の言語材料に基づいて構成される。例えば、それは以下の方法で構成することができる。

（１）例えば「ｗ１ｗ２，ｗ３。ｗ４」などの訓練テキストを例に取る。訓練テキストの擬似句読点を識別して、

ｗ１「無句読点」ｗ２コンマｗ３終止符ｗ４

10

＊は、以下の手順として認識することもできる。

【００４４】前記第二シーケンスの全ての擬似雑音ｄおよびその文脈ｃについて、第二音響モデル（ＡＭ２）および第二言語モデル（ＬＭ２）を使用することによって、

M^{ML}＝ａｒｇｍａｘ　ｍ：ＡＭ２（ｄ，ｍ）＊ＬＭ２（ｍ，ｃ）

となるように最適擬似句読点M^{ML}を見つける。ここで、ｍは句読点であり、

ＡＭ２（ｄ，ｍ）＝Ｐ（ｄ｜ｍ）

ＬＭ２（ｍ，ｃ）＝Ｐ（ｍ｜ｃ）

である。

【００４５】

を得る。

（２）訓練者は、句読点を読み上げずにテキスト「ｗ１ｗ２，ｗ３。ｗ４」を読み上げる。

（３）第一音響モデルＡＭ１および第一言語モデルＬＭ１を使用して、訓練者の音声訓練を復号する。上記テキストには句読点があるので、訓練者は読み上げるときに、特定の読み上げ様式を表現する。ｗ１とｗ２の間には句読点が無く、これらは連続的に読み上げられる。ｗ２を読み上げた後、訓練者はコンマに遭遇し、短時間休止し、吸息するかもしれない。次に訓練者はｗ３を読み上げ、沈黙する（終止符のため）。最後にｗ４を読み上げる。例えば、復号された出力は次のようになる。

ｗ１ｗ２吸息ｗ３沈黙ｗ４

（４）復号された出力に雑音をマークする。上記の例では、次のようになる。

ｗ１「無音」ｗ２吸息ｗ３沈黙ｗ４

（５）擬似句読点ｍと対応する擬似雑音ｄを照合する。

（「無句読点」，「無音」）

（コンマ，吸息）

（終止符，沈黙）

【００５２】擬似句読点ｍの種類および擬似雑音の種類について、対（ｍ，ｄ）と呼ばれる対応する関係がある。対（ｍ，ｄ）の数はｃ（ｍ，ｄ）と表される。訓練テキスト、つまり言語材料は、訓練者と同様に、様々な擬似句読点および普通の人々の話し方を充分に網羅する必要がある。したがって、ｃ（ｍ、ｄ）は一般に１より多い。

（６）Ｐ（ｄ｜ｍ）は大まかにｃ（ｍ，ｄ）／ｃ（ｍ）で推定される。ここでｃ（ｍ）は、全ての擬似雑音ｄ’の対応するｃ（ｍ，ｄ’）の合計である。

【００５３】上記は、第二音響モデルＡＭ２を構成する方法である。いうまでもなく、他の方法を使用して、同じ機能を持つ音響モデルＡＭ２を構成することができる。

11

【００５４】図２および図３に関連して以上で説明した句読点を自動的に生成するための装置および方法では、ユーザが句読点を声に出して言う必要も、ユーザが句読点の位置を指示する必要も無い。しかし、様々なユーザが様々な話し方をするので、擬似雑音を句読点の位置を指示するための条件の１つとして使用する場合、何らかのエラーが生じるはずである。

【００５５】以下で述べる第二実施形態では、口述中に句読点が必要なときに、口述と同時にユーザが明瞭な指示を与えることが必要である。そうした明瞭な指示は、例えば、マウスボタンまたは特定のハードウェアをクリックすることによって実現される。したがって、擬似雑音を使用しないので、第一音響モデルＡＭ１の擬似雑音部および第二音響モデルＡＭ２は除去される。実現が容易になり、より高い精度を得られる。しかし、ユーザにとっては、第一実施形態の場合ほど操作が簡便ではない。

【００５６】図４に示すように、本発明に従って句読点を自動的に生成する方法の第二実施形態は、以下の段階を含む。

【００５７】ステップＳ４１で、音声認識手順が開始する。このステップでは、文脈ｃなどの内部変数は全て空にされる。

【００５８】ステップＳ４２で、ユーザの音声が検出される。ステップＳ４３で、通常の音響モデルＡＭおよび言語モデルＬＭを使用することによって、ユーザの音声が語に復号される。

【００５９】ステップＳ４５で、口述中にユーザによって指示される句読点の位置が識別される。

【００６０】ステップＳ４７で、次のように第二言語モデルＬＭ２を使用することによって、最適擬似句読点M^{ML}が見つけ出される。

M^{ML}＝ａｒｇｍａｘ　ｍ：ＬＭ２（ｍ，ｃ）

ここで、ｍは句読点であり、

ＬＭ２（ｍ，ｃ）＝Ｐ（ｍ｜ｃ）である。

【００６１】ステップＳ４８で、M^{ML}は自動的に生成された句読点として選ばれ、現在の文脈ｃが更新される。

【００６２】ステップＳ４９で、連続音声認識を終了するかどうかが判断される。終了しない場合には、手順はＳ４２にジャンプする。そうでなければ、ステップＳ４１０に進む。

【００６３】ステップＳ４１０で、認識された語および自動的に生成された句読点が出力される。このステップで、擬似句読点を実句読点に置換することができる。

【００６４】ステップＳ４１１で、手順は終了する。

【００６５】次に第三実施形態について説明する。これは、機能的に第一実施形態と第二実施形態の中間的な形態である。第三実施形態は、ユーザが口述中に句読点が必要なときに明確な指示を与えなければならないが、ユーザは、任意の検出可能な雑音を発生するために例えば「唇を打つ音」など特殊な音を発するか、または句読点を指示するために、物理的な移動を行うことなく、意図的に沈黙するだけでよいという点が、第二実施形態とは異なる。この方法により、ユーザはより便利に流暢に話すことができる。第三実施形態は、口述中に句読点の位置で特殊な音を発生するので、自然な雑音と句読点を指示する音との間の相違がいっそう明瞭になるという点で、第一実施形態とは異なる。第二音響モデルＡＭ２を構成するときの訓練者の要件は同じである。第三実施形態は、第一実施形態より高い精度が得られることが、実践的に証明されている。

12

【００６６】本発明の方法は、必ずしも後処理に限定されるものではない。つまり、文全体の復号が完了した後で句読点を自動的に生成する必要はなく、実時間で生成することができる。つまり、文脈を形成する語が復号されるや否や、文脈に従って句読点を自動的に生成することができる。

【図面の簡単な説明】

【図１】従来技術の連続音声認識システムの構成の略図である。

【図２】本発明に係る連続音声認識システムで句読点を自動的に生成するための装置の第一実施形態の一般的構造の略図である。

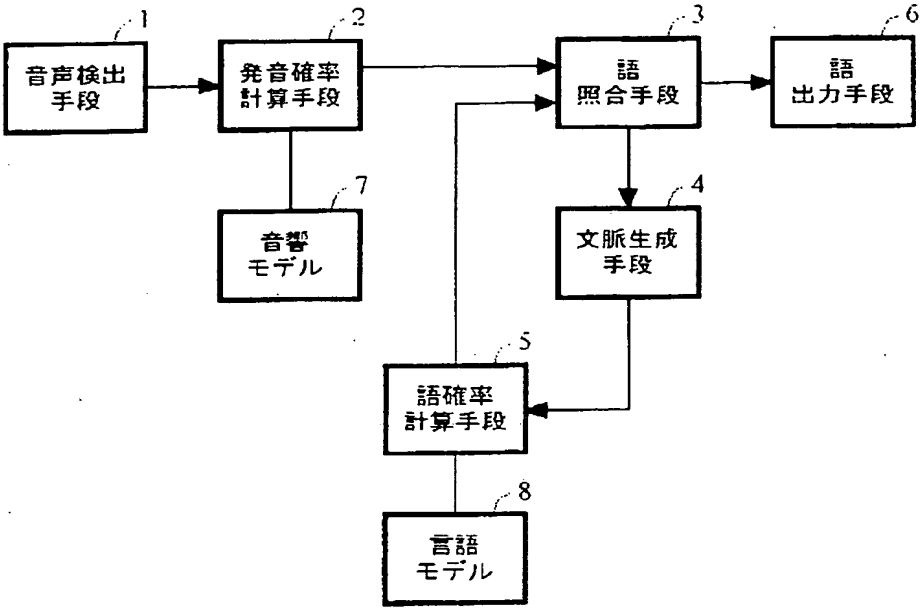
【図３】本発明に係る連続音声認識で句読点を自動的に生成するための方法の第一実施形態の一般流れ図である。

【図４】本発明に係る連続音声認識で句読点を自動的に生成するための方法の第二実施形態の一般流れ図である。

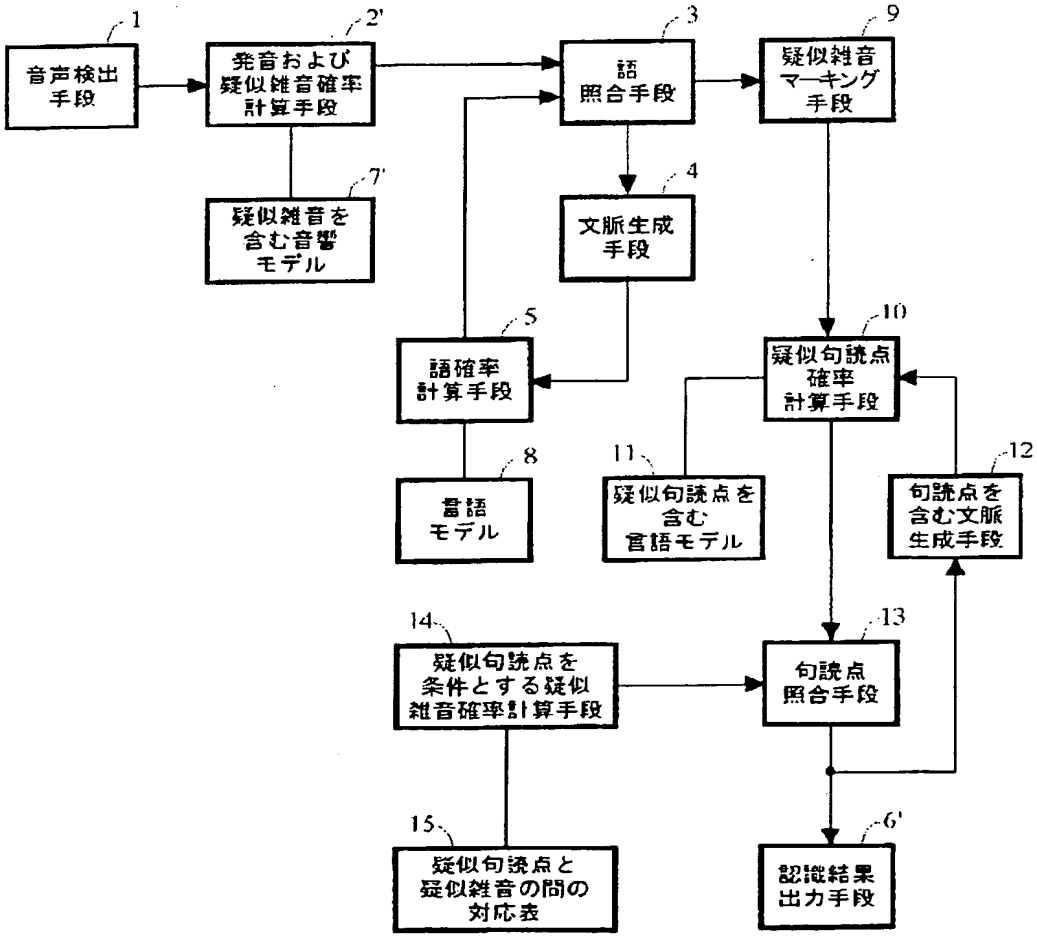
【符号の説明】

１　音声検出手段
２’　発音および擬似雑音確率計算手段
３　語照合手段
４　文脈生成手段
５　語確率計算手段
６’　認識結果出力手段
７’　擬似雑音を含む音響モデル
８　言語モデル
９　擬似雑音マーキング手段
１０　擬似句読点確率計算手段
１１　擬似句読点を含む言語モデル
１２　句読点を含む文脈生成手段
１３　句読点照合手段
１４　擬似雑音確率計算手段
１５　対応表

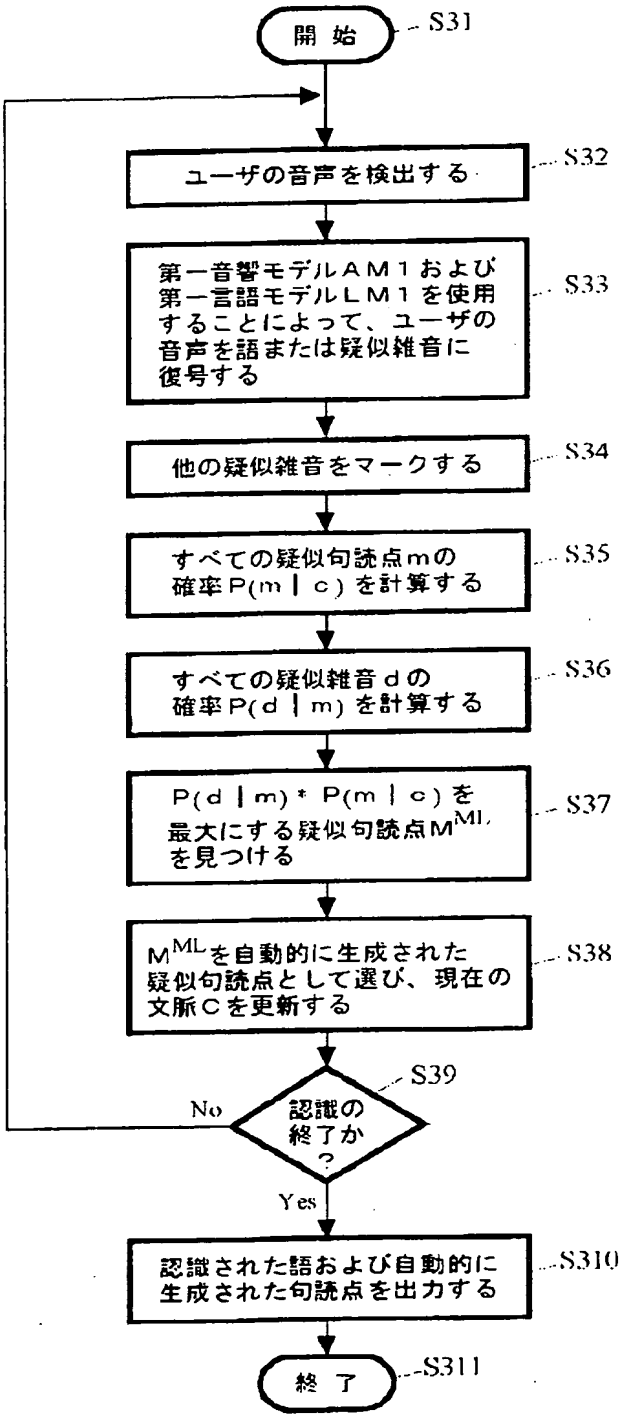
【図１】


1
音声検出手段
2
発音確率計算手段
3
語照合手段
6
語出力手段
7
音響モデル
4
文脈生成手段
5
語確率計算手段
8
言語モデル

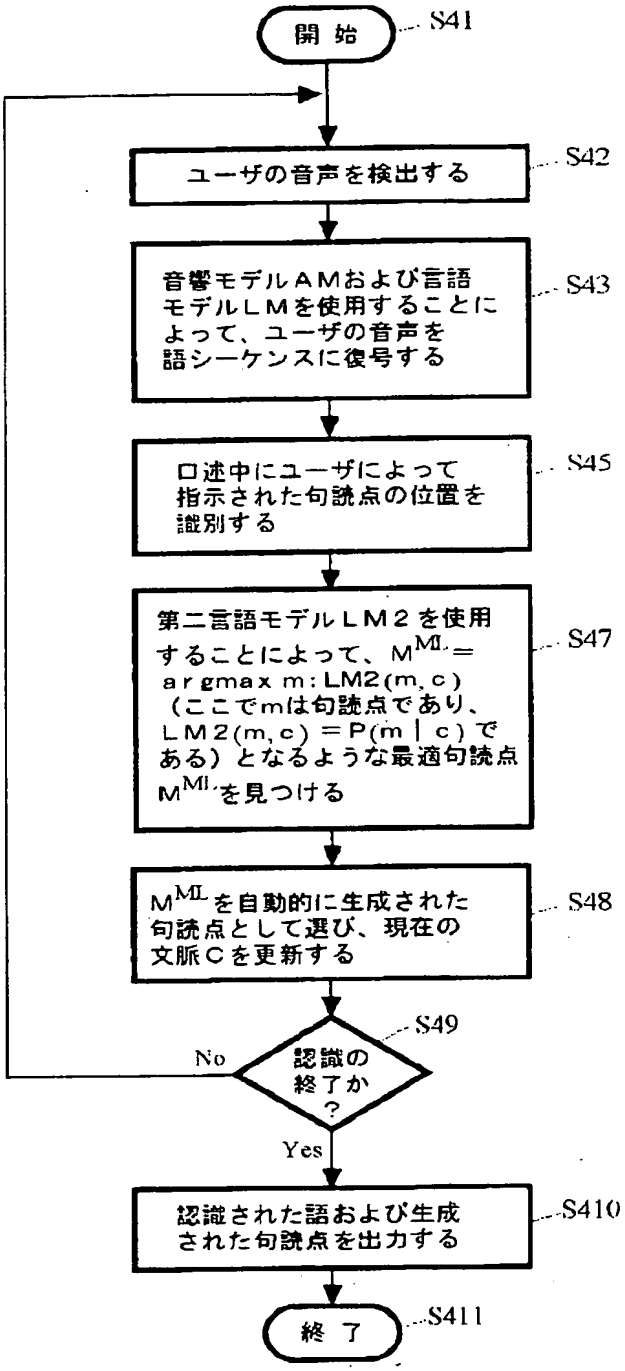
【図２】


1
音声検出手段
2'
発音および疑似雑音確率計算手段
3
語照合手段
9
疑似雑音マーキング手段
7'
疑似雑音を含む音響モデル
4
文脈生成手段
5
語確率計算手段
10
疑似句読点確率計算手段
8
言語モデル
11
疑似句読点を含む言語モデル
12
句読点を含む文脈生成手段
14
疑似句読点を条件とする疑似雑音確率計算手段
13
句読点照合手段
15
疑似句読点と疑似雑音の間の対応表
6'
認識結果出力手段

【図３】


開 始
S31
ユーザの音声を検出する
S32
第一音響モデルＡＭ１および第一言語モデルＬＭ１を使用することによって、ユーザの音声を語または疑似雑音に復号する
S33
他の疑似雑音をマークする
S34
すべての疑似句読点ｍの確率Ｐ(ｍ｜ｃ)を計算する
S35
すべての疑似雑音ｄの確率Ｐ(ｄ｜ｍ)を計算する
S36
Ｐ(ｄ｜ｍ)＊Ｐ(ｍ｜ｃ)を最大にする疑似句読点M^{ML}を見つける
S37
M^{ML}を自動的に生成された疑似句読点として選び、現在の文脈Ｃを更新する
S38
No
認識の終了か？
S39
Yes
認識された語および自動的に生成された句読点を出力する
S310
終 了
S311

【図４】


開 始
S41
ユーザの音声を検出する
S42
音響モデルＡＭおよび言語モデルＬＭを使用することによって、ユーザの音声を語シーケンスに復号する
S43
口述中にユーザによって指示された句読点の位置を識別する
S45
第二言語モデルＬＭ２を使用することによって、M^{ML}＝ａｒｇｍａｘ ｍ:ＬＭ2(m,c)（ここでｍは句読点であり、ＬＭ2(m,c)＝Ｐ(ｍ｜ｃ)である）となるような最適句読点M^{ML}を見つける
S47
M^{ML}を自動的に生成された句読点として選び、現在の文脈Ｃを更新する
S48
No
認識の終了か？
S49
Yes
認識された語および生成された句読点を出力する
S410
終 了
S411

フロントページの続き

(72)発明者　ドナルド・ティー・タン
中華人民共和国　北京市朝陽区　アジア大会村　フイユアン・インターナショナルアル・アパートメンツ　ディー棟　アパートメント1708

(72)発明者　シャオ・チン・チュー
中華人民共和国　北京市海淀区　アンニンリー　15－4－402

(72)発明者　リー・チン・シェン
中華人民共和国　北京市海淀区　シャンティートンリー　セクション2　3－3－102